PhytoKeys

PhytoKeys 244: 57–76 (2024)
DOI: 10.3897/phytokeys.244.120914

Research Article

Genetic and morphological evidence support the specific status of the endemic *Erica andevalensis* (Ericales, Ericaceae)

Iván Rodríguez-Buján[1,2], Pilar Díaz-Tapia[3], Jaime Fagúndez[1,2]

1 *Universidade da Coruña, BIOCOST research group, Centro Interdisciplinar de Química e Bioloxía (CICA), Rúa As Carballeiras, 15071, A Coruña, Spain*
2 *Universidade da Coruña, Facultade de Ciencias, Departamento de Bioloxía, 15071, A Coruña, Spain*
3 *Universidade de Santiago de Compostela, Facultade de Bioloxía, Departamento de Botánica, 15072 Santiago de Compostela, Spain*
Corresponding author: Jaime Fagúndez (jaime.fagundez@udc.es)

Abstract

Assessing the taxonomic status of closely related taxa is crucial in plant systematics and can have important implications for conservation and human plant use. *Erica andevalensis* Cabezudo & Rivera is a metallophyte endemic species from highly metal-polluted soils of SW Iberian Peninsula, an area with a mining history going back more than 5,000 years. *Erica andevalensis* is closely related to *Erica mackayana* Bab., a northern Iberian species also present in western Ireland. The status of *E. andevalensis* as a species or subspecies subordinated to *E. mackayana* is subject to debate. Here, we assessed the genetic and phenotypic relationship between both species, including the population structure of *E. andevalensis*. We used high throughput sequencing to determine genome-wide Single Nucleotide Polymorphisms (SNPs), and morphometric analyses from 35 reproductive and vegetative traits. The morphological analysis showed at least eight characters that can discriminate the two species, from which ovary hairiness and the size of leaf glandular hairs were the most informative. Genetic analyses showed that each species formed a monophyletic cluster with full support, separated by an interspecific genetic distance >4-fold higher than intra-specific distance. Population genetic analyses of *E. andevalensis* shows that populations are highly structured, with the Portuguese one as the most isolated and less variable. These results support the recognition of *E. andevalensis* as a distinct species with a highly constrained ecological requirements and a narrow geographic distribution, but with a limited gene flow between populations. We discuss the implications of these outcomes in conservation policies and potential uses of *E. andevalensis* such as decontamination of polluted soils.

Key words: Endemic species, *Erica*, metallophyte, population structure, species delimitation, systematics

OPEN ACCESS

This article is part of:
Systematics, natural history, and conservation of* Erica *(Ericaceae)
Edited by Michael Pirie, Félix Forest, Timo van der Niet, Jaime Fagúndez, Seth Musker, Fernando Ojeda, Anina Coetzee, E. Charles Nelson

Academic editor: Michael Pirie
Received: 14 February 2024
Accepted: 23 May 2024
Published: 5 July 2024

Citation: Rodríguez-Buján I, Díaz-Tapia P, Fagúndez J (2024) Genetic and morphological evidence support the specific status of the endemic *Erica andevalensis* (Ericales, Ericaceae). PhytoKeys 244: 57–76. https://doi.org/10.3897/phytokeys.244.120914

Copyright: © Iván Rodríguez-Buján et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

The species is the fundamental unit in taxonomy and systematics of living organisms (De Queiroz 2005). This perspective extends to all fields of biology, leading to considerable scientific effort being dedicated to this taxonomic category. Beyond a scientific perspective, species delimitation is relevant in conservation planning and biodiversity assessment of protected areas (Coates et al. 2018; Galtier 2019).

In plant taxonomy, morphological characters have historically been the primary source of evidence for delimitating species (Whittemore 1993; Soltis and Soltis 2012). However, the more recent application of DNA data to assess species diversity has significantly impacted the field of taxonomy. Ideally, species delimitation would be informed by a comprehensive understanding of variability and population structure, as well as knowledge of the connections among closely related lineages (Galtier 2019). Since most species exhibit geographic variation, it is important to obtain a good representation of the distribution area to distinguish true discontinuities, which could imply lineage separation, from within-lineage variation (De Queiroz 2007). In lineages with incipient separation, considering other lines of evidence (e. g. ecological, morphological, and/or phenological) in addition to molecular data provides the basis for a more robust integrative taxonomy (Pante et al. 2015).

High throughput sequencing (HTS) is the most significant recent advance in molecular techniques, as it greatly facilitates the generation of large amounts of DNA data. Restriction site-associated DNA (RAD-seq) and similar approaches involve enzymatic fragmentation of genomic DNA coupled with HTS to determine large numbers of molecular markers with genome-wide coverage (Andrews et al. 2016). The enhanced resolution provided by these techniques has been applied to disentangle complex relationships between closely related species as well as to study the genetic structure of populations (e.g. Pante et al. 2015; Feng et al. 2018).

Erica is among the largest genera of seed plants, with 851 accepted species (Elliott et al. 2024; Oliver et al. 2024; WFO 2024), most of them concentrated in the Cape Floristic Region (Manning and Goldblatt 2012). However, the genus most probably originated in the Palearctic region, where only 23 species live today (Mugrabi de Kuppler et al. 2015). Despite the relatively low number of *Erica* species in Europe, they hold a large phylogenetic diversity compared to the species-rich Cape floristic region (Pirie et al. 2024). *Erica* also plays an important ecological role as dominant species in different European habitats (Fagúndez 2013). European heathlands are facing many threats, and their habitats are decreasing, resulting in the isolation of populations and local extinctions (Fagúndez 2013). Understanding the genetic diversity and population structure of these species becomes essential to assess the conservation status of their populations and infer their future survival in a context of global change (Frankham 2010).

Erica andevalensis Cabezudo and Rivera, occurs in a restricted area in the south-west of the Iberian Peninsula, mainly in western Andalusia, Spain, and bordering Portugal (Cabezudo and Rivera 1980; Buira et al. 2017). It colonizes wet soils within the Iberian Pyrite Belt, characterized by high levels of sulfur and heavy metals. It frequently forms nearly monospecific communities along polluted riversides and in abandoned mines (Aparicio 1999; Fig. 1). As a restricted metallophyte endemic species, *E. andevalensis* is considered a threatened species (BOJA 2012; Carapeto et al. 2020).

Erica andevalensis is closely related to *E. mackayana* and both species are morphologically similar. However, they have disjunct distributions, as *E. mackayana* is restricted to the northern Iberian Peninsula (Mugrabi de Kuppler et al. 2015; Fagúndez and Díaz-Tapia 2023). Despite morphological similarities, both species can be distinguished by the pattern and size of glandular hairs and the presence/absence of other small non glandular hairs (Cabezudo and Rivera

Figure 1. A Flowering stems of *Erica andevalensis* **B** Tinto river shores in Huelva, SW Spain, the habitat of *E. andevalensis* **C** São Domingos, Portugal, abandoned mining area with *E. andevalensis* in front **D** *Erica andevalensis* in the shores of Odiel river **E** flowering stems of *Erica mackayana* **F** wet heathland dominated by *E. mackayana* in Galicia, NW Spain.

1980). However, some variability was found in these characters and both species can lack glandular hairs. As a result, *E. andevalensis* has been considered a subspecies under *E. mackayana* (McClintock and Nelson 1989), although Nelson (2012) subsequently treated *E. andevalensis* as a species in its own right. Additional research has revealed further characters that distinguish both taxa, as they differ in seed morphology (Fagúndez and Izco 2004). Seed characters have been proven useful in delimitating *Erica* species (e.g. Szkudlarz 2009; Fagúndez and Izco 2010). Also, Bayer (1993) described small hairs in the ovary of some individuals of *E. andevalensis* that were never observed in *E. mackayana*. The

discussion on the taxonomic rank of *E. andevalensis* continues (see Nelson and Small 2000; BOJA 2012; Carapeto et al. 2020; GBIF 2023; WFO 2024).

In this study, we aim to reassess the taxonomic identity of *E. andevalensis* as a distinct species from *E. mackayana* combining molecular and morphological data. By studying the whole geographic range of both species, we aim to identify morphological traits that might reliably distinguish *E. andevalensis* from *E. mackayana*. In addition, we aim to study the population structure of *E. andevalensis*, to understand the level of isolation and gene flow among populations and provide better guidelines for conservation.

Material and methods

Plant material

We sampled 38 plants from four populations of *E. andevalensis* in August 2021. Seventy-one plants from ten populations of *E. mackayana* analysed in Fagúndez and Díaz-Tapia (2023) were also included in the analyses (Fig. 2, Suppl. material 1: table S1). Additionally, one *E. tetralix*, the closest species to the *E. mackayana-andevalensis* clade (Mugrabi de Kuppler et al. 2015), was included as outgroup in the phylogenetic analyses.

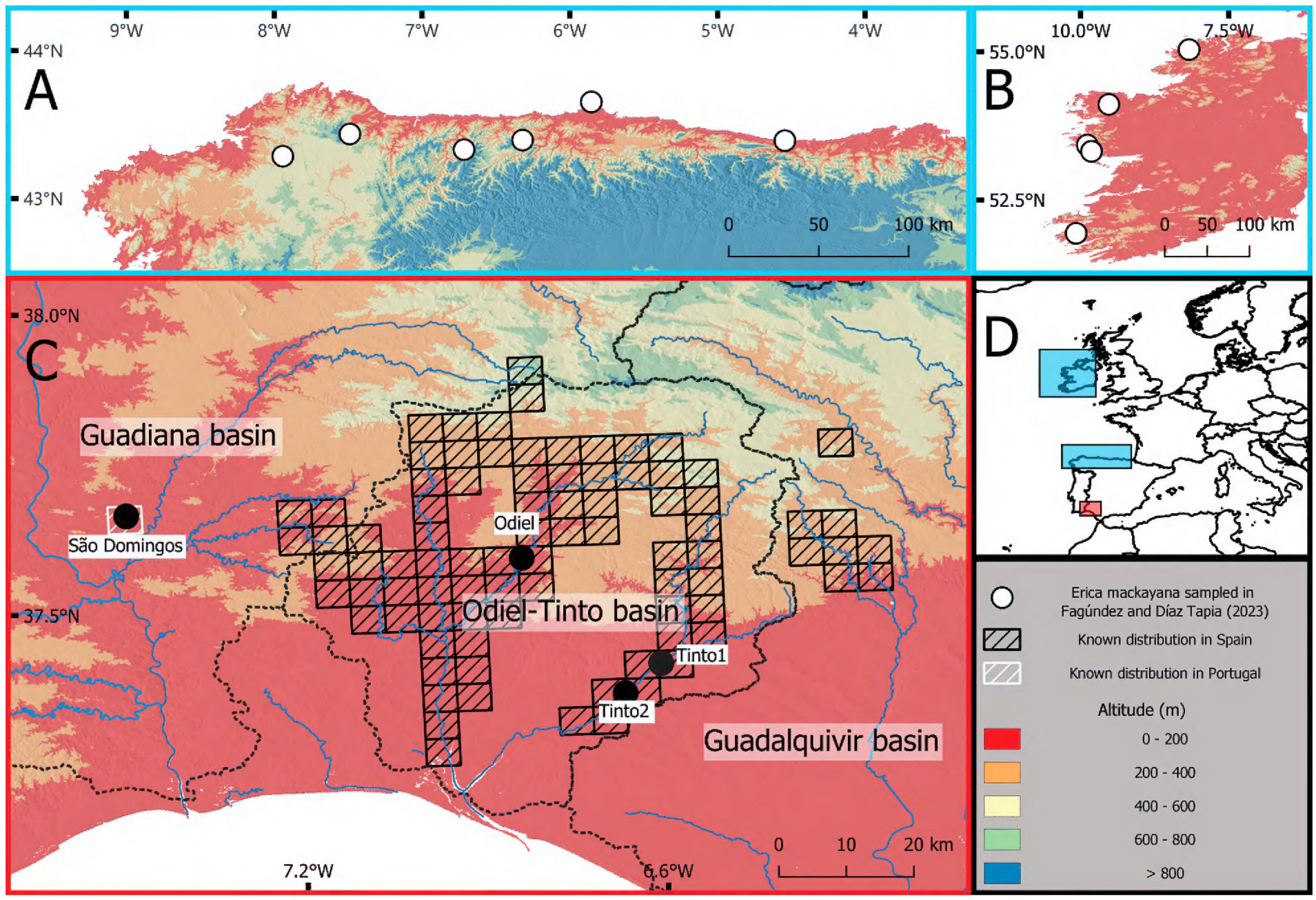

Figure 2. Distribution of the sampled populations of *E. andevalensis* and *E. mackayana* **A** *Erica mackayana* populations (as in Fagúndez and Díaz-Tapia 2023) sampled in NW Iberian Peninsula **B** *Erica mackayana* populations (as in Fagúndez and Díaz-Tapia 2023) sampled in Ireland **C** *Erica andevalensis* populations sampled. The distribution of *E. andevalensis* is represented in a 5×5 UTM grid from https://www.ideandalucia.es/catalogo/inspire/srv/api/records/625a6b54-bfc1-4589-8571-b4503bf262c2 (Spain) and Carapeto et al. (2020) (Portugal) **D** location of the three studied regions in Europe.

A subsample of each specimen was dried in silica gel, and the remaining material was prepared for morphological studies. All herbarium vouchers were deposited in SANT herbarium.

Morphometric data and analyses

Each sample was mounted in a herbarium sheet and scanned at a minimum resolution of 1000 dpi. In addition, a minimum of two fresh flowers per specimen were dissected to measure ovaries and anthers. Pictures of ovaries and anthers were taken using an OLYMPUS C3040-ADU for subsequent digital analysis. For leaf micro-characters, we used an optical microscope at 40–100 magnifications. All images were analyzed using ImageJ software (Schneider et al. 2012).

The final dataset for morphological analyses comprises 71 individuals of *E. mackayana* and 38 of *E. andevalensis*. In the dataset, each trait was represented by one value per specimen, thus the arithmetic mean was calculated for multiple measures (Suppl. material 1: table S2). We measured a total of 35 traits including 24 quantitative, 8 ordinal and 3 binary (Fig. 3, Suppl. material 1: table S2). Each trait was measured once or several times per sample. We tested for differences between the means in the two species for each variable using a non-transformed dataset. For quantitative variables, we applied t-tests or Wilcoxon rank-sum tests, depending on whether the variables met normality and homocedasticity assumptions. Ordinal variables were compared using Chi-squared tests. In the case of binary variables (LNR, L5 and LS2G), the Fisher's exact test was conducted (Suppl. material 1: table S3).

We performed a principal component analysis (PCA) with all samples and traits. Since each trait was on a different scale, the prep function from the PCAMETHODS package (Stacklies et al. 2007) was used to normalize and center all variables. Missing data in the final matrix represented 12.08%. To overcome this issue, we used a K-Nearest-Neighbor approach from DMWR2 package (Torgo 2016). Subsequently, the FACTOEXTRA package was used to perform the PCA (Kassambara and Mundt 2020). Next steps were carried out using the MORPHOTOOLS2 package (Šlenker et al. 2022). All statistical analysis were conducted in R (R Core Team 2020).

Molecular data

We used genomic data obtained from Fagúndez and Díaz-Tapia (2023) for *E. mackayana*, while genomic data of *E. andevalensis* was newly determined. One sample of *E. tetralix* from Fagúndez and Díaz-Tapia (2023) was included in the dataset to be used as the outgroup in phylogenetic analyses. DNA extraction, library preparation and sequencing followed Fagúndez and Díaz-Tapia (2023).

Nextera adaptors from sample reads were trimmed using BBDUK (BBMAP TOOLS v.38.79, http://sourceforge.net/projects/bbmap/) with the following parameters: ktrim = r, k = 17, hdist = 1, mink = 8, minlen = 100, ow = t, qtrim = r, trimq = 10. Quality of filter reads was checked using FastQC (v.0.11.9) (Andrews 2010). Denovo_map.pl pipeline from STACKS software v2.64 (https://catchenlab.life.illinois.edu/stacks/) was employed to denovo assembly and SNP calling.

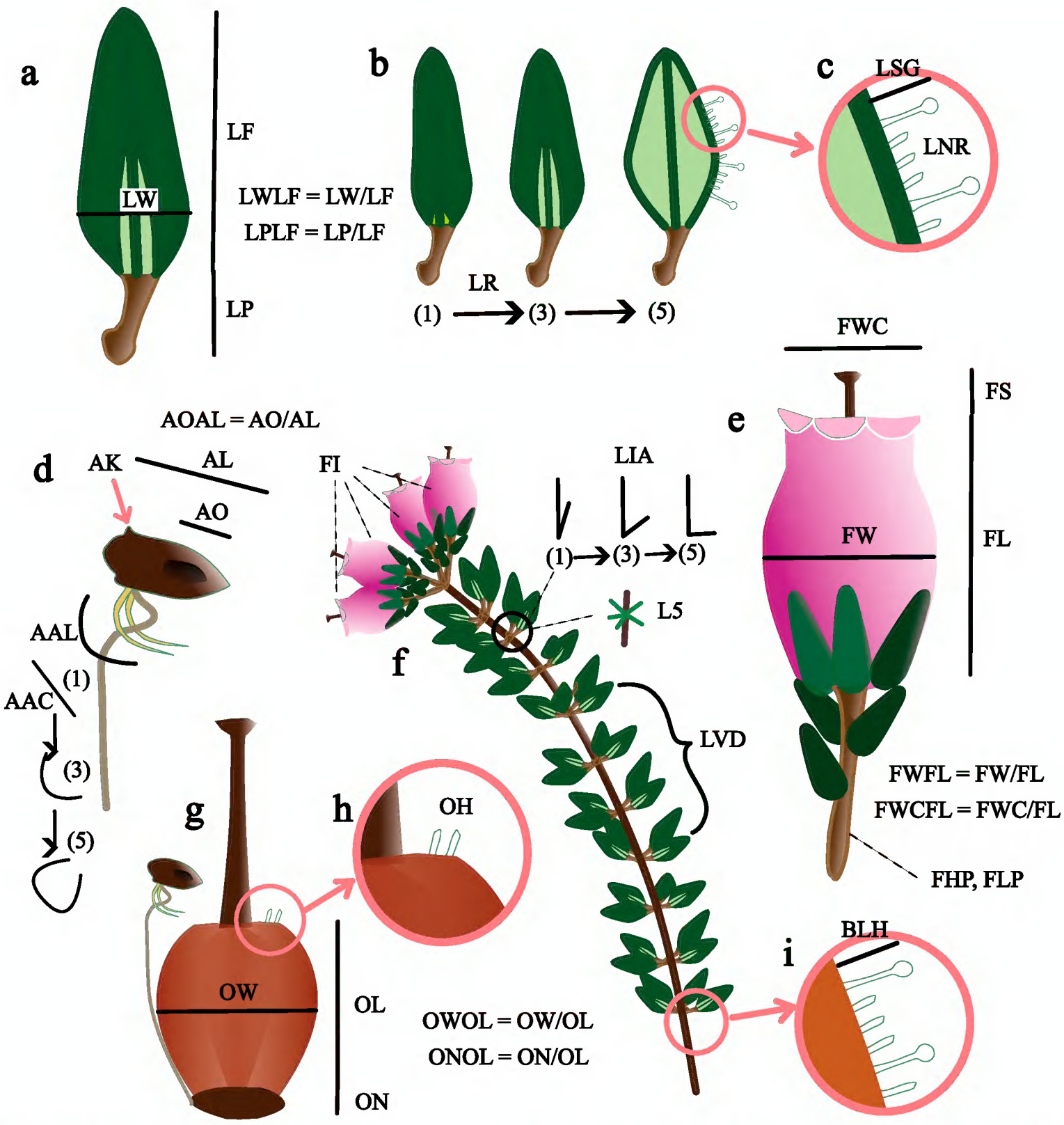

Figure 3. Schematic representation of the morphological traits measured in the samples of *Erica andevalensis* and *E. mackayana*. Leaf traits (**a**) (LF = Leaf length; LW = Leaf width; LP = Petiole length), variation on leaf morphology (**b**) (LR = Leaf rolling degree), and indumentum (**c**) (LSG = length of glandular hairs; LNR = presence of non-glandular hairs). Stamen traits (**d**) (AL = Anther length; AO = Anther pore; AAL = Anther appendix length; AAC = Anther appendix curvature; AK = Anther knob), flower morphology (**e**) (FL = Corolla length; FW = Corolla width; FWC = Corolla opening; FS = Style exertion; FHP = Pedicel hairiness, FLP = Pedicel length), and arrangement of a flowering branch (**f**) (FI = Flowers per inflorescence; LIA = Leaf insertion angle, L5 = 5nate leaves, LVD = Density of whorls). The ovary (**g**) (OL = Ovary length; OW = Ovary width; ON = Nectary size) and detail of ovary surface (**h**) (OH = Ovary hairiness) and stem (**i**) (BLH = Branch longest hair). Coding, measures, and additional information in Suppl. material 1: table S2.

To select the optimal parameters, a preliminary optimization step following the 80% rule (Paris et al. 2017) was conducted. Thirty percent of the samples with a minimum of two per population were selected. Reads with different lengths were allowed using –force-diff-len on ustacks module. We constructed 12 different catalogs, varying the -m (Minimum number of raw reads required to form a stack (a putative allele)), -M (Number of mismatches allowed between stacks (putative alleles) to merge them into a putative locus) and -n (Number of mismatches allowed between stacks (putative loci) during construction of the catalog) parameters. For all probes, -M = -n, but no -m, which ranged between 2 and 6. After optimization process, we created a new catalog choosing m = 4, M = 3, n = 3. Using a populations' module, the dataset was exported to variant call format (VCF).

Phylogenetic analysis

VCF file from STACKS was filtered with VCFTOOLS retaining the SNPs with less than 60% of missing data and those that were separated by at least 200 bp to avoid linkage disequilibrium issues. Furthermore, only variable SNPs present in a minimum of 4 samples were accepted. The VCF file was exported as interleaved phyllip using TASSEL (Bradbury et al. 2007). IQTREE v2.1-2 was used to filter invariant sites (Minh et al. 2020). Final dataset consisted of 6004 bp. MODELFINDER from WEB-IQTREE (Kalyaanamoorthy et al. 2017) was used to select the optimal model of the dataset based on Bayesian Inference Criterion (BIC). Using this approach, the best model was TVM+F+ASC+G4. Then, the final tree was calculated with this model in IQ-TREE v2.1-2 with 1000 non-parametric bootstrap replicates.

Genetic distance analysis

To assess genetic distance between *E. andevalensis* and *E. mackayana*, we used Tamura-Nei distance (Tamura and Nei 1993) in MEGA11 (Tamura et al. 2021). We compared only populations with >2 individuals. Then, we calculated mean pairwise distances for population of each species and compared them with mean pairwise distance between both species.

Population structure analysis

A new dataset per putative species was created. VCFTOOLS was used to remove indels, as well as filter SNPs retaining those that were present in at least 80% of samples and minor alleles were present in at least two samples. Also, to avoid linked loci, SNPs at a distance less than 500 base pairs (bp) were excluded. Both datasets were exported to plink format. The function filter_data() from SAMBAR package was used with default thresholds to detect and erase possible paralogous loci (De Jong et al. 2021). Final datasets consisted of 36 samples with 4,234 SNPs of *Erica andevalensis* and 62 samples with 5,178 SNPs of *E. mackayana*.

Observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated with function filter_data() in SAMBAR, and nucleotide diversity was calculated with calc_diversity() of the same package. F_{IS} was calculated as $1-H_o/H_e$ as in Weir and Cockerham (1984).

We further analysed the dataset of *E. andevalensis*, including a F_{ST} pairwise comparison among populations, using calc_diversity() function in SAMBAR and Bayesian population assignment (BPA). BPA probabilities were calculated using the function find_structure() in SAMBAR package. The optimal number of clusters (K) was determined using the elbow method on cross-entropy scores.

Results

Morphometric analyses

Twenty-eight out of 35 morphological characters measured showed statistically significant differences between *Erica mackayana* and *E. andevalensis* (Suppl. material 1: table S3).

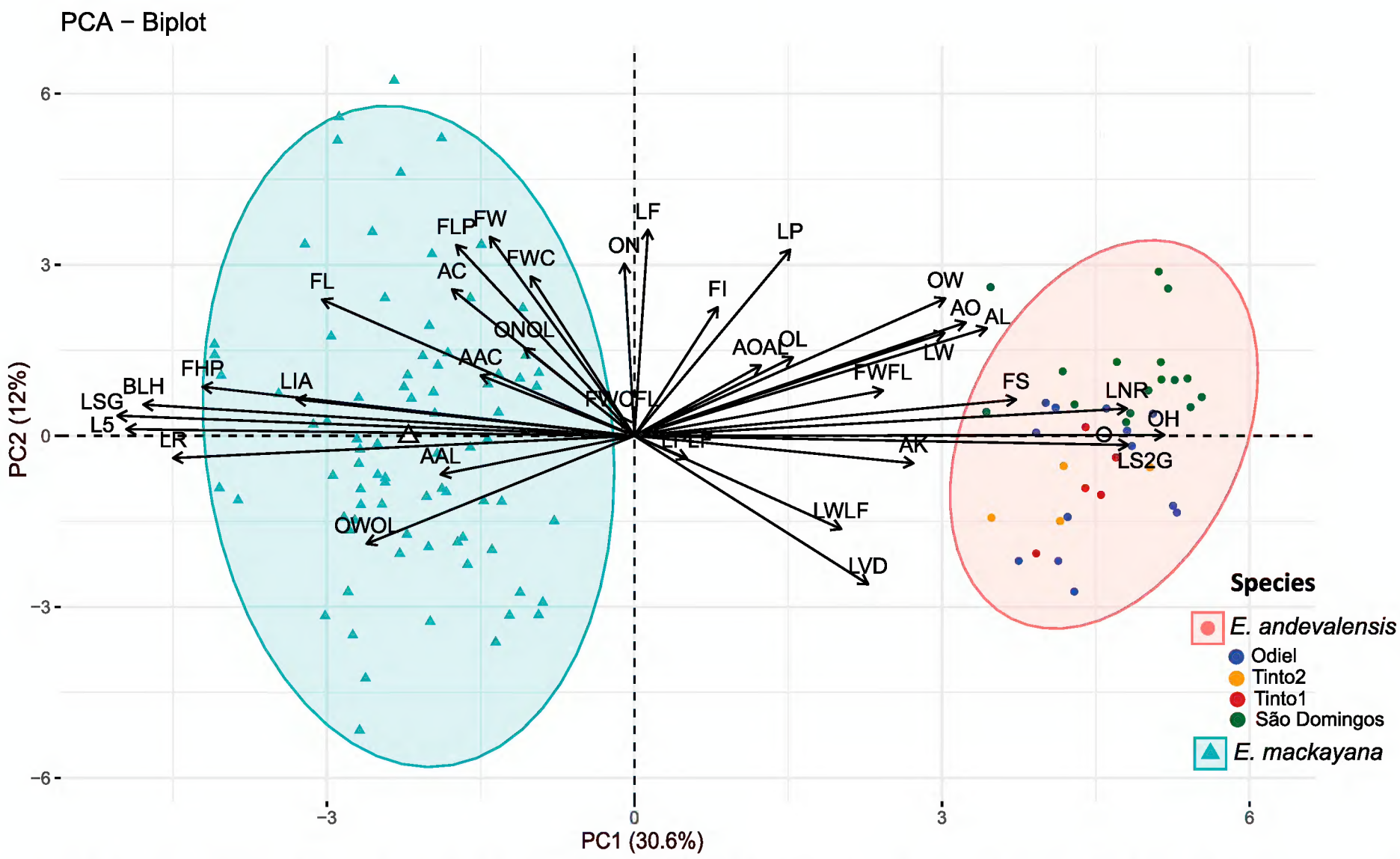

Figure 4. Biplot showing the first two principal components. Each ellipse represents the area that would encompass 95% of individuals assuming populations follow a normal distribution. White dots and triangles represent *E. andevalensis* and *E. mackayana* centroids respectively. *Erica andevalensis* populations are represented in different colors: Green (São Domingos), blue (Odiel), brown (Tinto1), yellow (Tinto2). Trait acronyms as in Fig. 3 and Suppl. material 1: table S2.

The first and second axes of the PCA explained 30.6% and 12% of the variance in the data, respectively. PC1 clearly delimitate the two species into two separated clusters (Fig. 4). PC2 depicts intra specific variation, with the species centroids close to 0 (-0.01 in *E. mackayana* and 0.02 in *E. andevalensis* respectively). PC2 can delimitate among *E. andevalensis* populations, particularly Spanish populations (Tinto1, Tinto2 and Odiel) populations vs São Domingos from Portugal. We identified the best diagnostic characters as those with a higher value on the first PC of the PCA, thus contributing to delimitate the two species, but lower values on the second PC, which represents variability within both taxa. Eight were selected as the best diagnostic characters including ovary hairiness, length and arrangement pattern of glandular hairs in leaves, length of glandular hairs on the stem, presence of non-glandular hairs in leaves, inflorescence hairiness, leaf rolling degree and number of leaves per whorl (Table 1, Fig. 5). The contribution of the remaining variables is provided in Suppl. material 1: table S4.

Phylogenetic analysis

The phylogenetic tree placed the samples of the two species in two distinct clades that received full support (Fig. 6). Within the *E. andevalensis* clade, the samples from Tinto 1 and São Domingos formed two fully supported clades. The samples from Tinto 2 were placed as sister to Tinto 1, but this relationship was unsupported. Most samples from Odiel were placed in a moderately supported clade, except two individuals of which one was sister to the clade formed plants from Tinto 1 and 2 and the other was sister to the clade formed by all the other samples.

Table 1. Selection of traits that most contribute to the two first PCA components. Percentage of variability within traits explained by the first two components. For each trait average values for each species are provided. Significance of correspondent mean differences tests is highlighted with *** if p-value <0.001. OH = Ovary hairiness; LSG = Length of glandular hairs; L5 = Presence of 5nate whorls; LS2G = Presence of more than two rows of glandular hairs; LNR = Presence of non-glandular hairs; BLH = Longest hair length in stem; LR = Leaf rolling degree; FHP = Pedicel hairiness.

Trait	Units/coding	Contribution to PC1	Contribution to PC2	*E. mackayana* mean	*E. andevalensis* mean	Significance
OH	0–5	8.584	0.000	0	1.083	***
LSG	µm	8.173	0.103	562.025	136.448	***
L5	0–1	7.890	0.011	0.958	0	***
LS2G	0–1	7.452	0.021	0	0.773	***
LNR	0–1	7.412	0.187	0	0.839	***
BLH	µm	7.369	0.246	0.986	0.333	***
LR	1–5	6.506	0.125	2.937	1	***
FHP	1–5	5.703	0.607	2.500	1.105	***

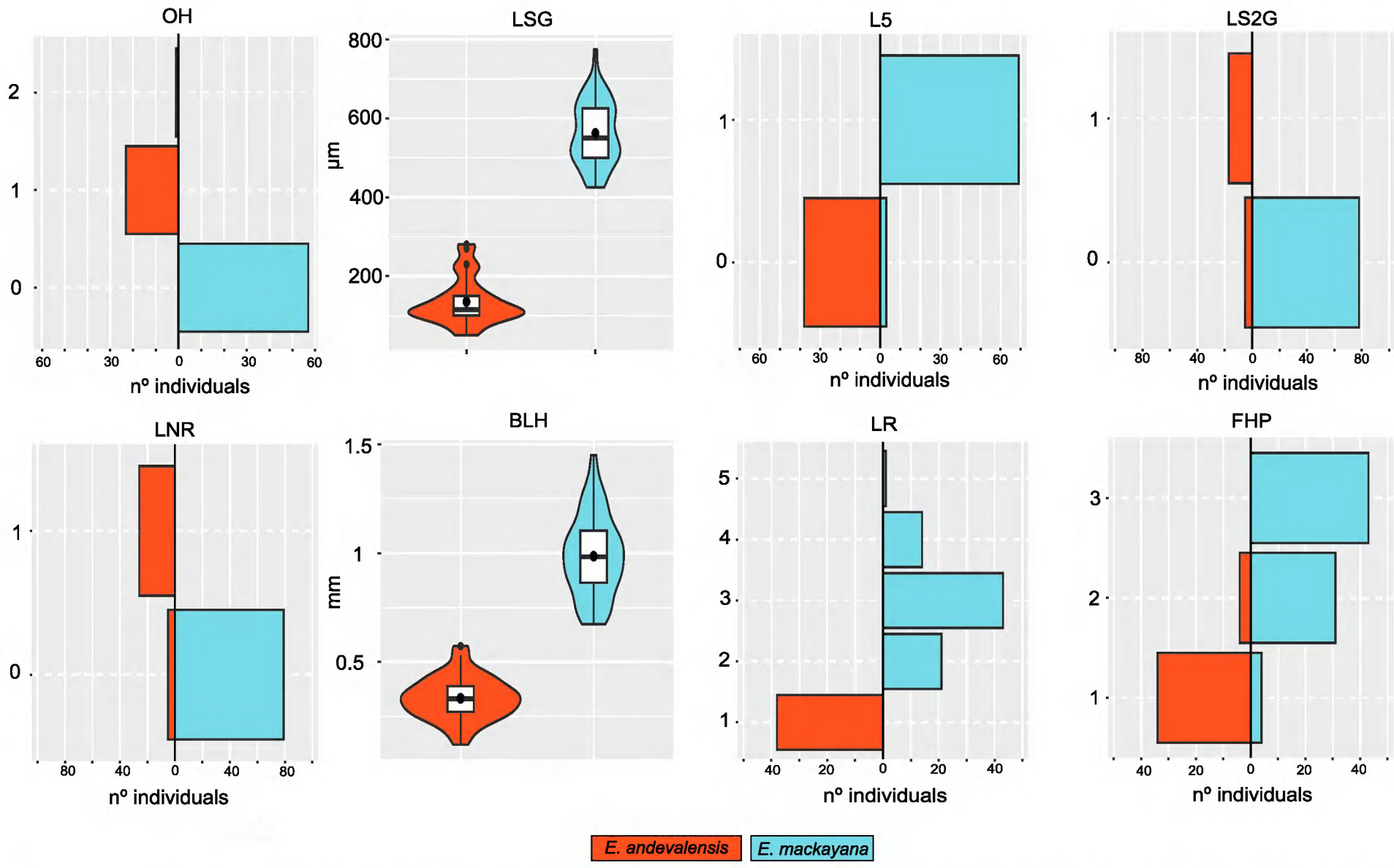

Figure 5. Boxplots and histograms of the traits with the highest contribution to PC1 of the PCA as in Table 1. Red corresponds to *E. andevalensis*, blue to *E. mackayana*. Black dots represent the mean. OH = Ovary hairiness; LSG = Length of glandular hairs; L5 = Presence of 5nate whorls; LS2G = Presence of more than 2 rows of glandular hairs; LNR = Presence of non-glandular hairs; BLH = Longest hair length in stem; LR = Leaf rolling degree; FHP = Pedicel hairiness.

Genetic distance

Nei's genetic distance among *E. mackayana* populations was 0.280 in average, while distance among *E. andevalensis* populations was 0.268 (Suppl. material 1: table S5). Mean Nei's genetic distance between both species was 1.146. Therefore, the inter-specific average distance between both species was >4 times greater than intra-specific.

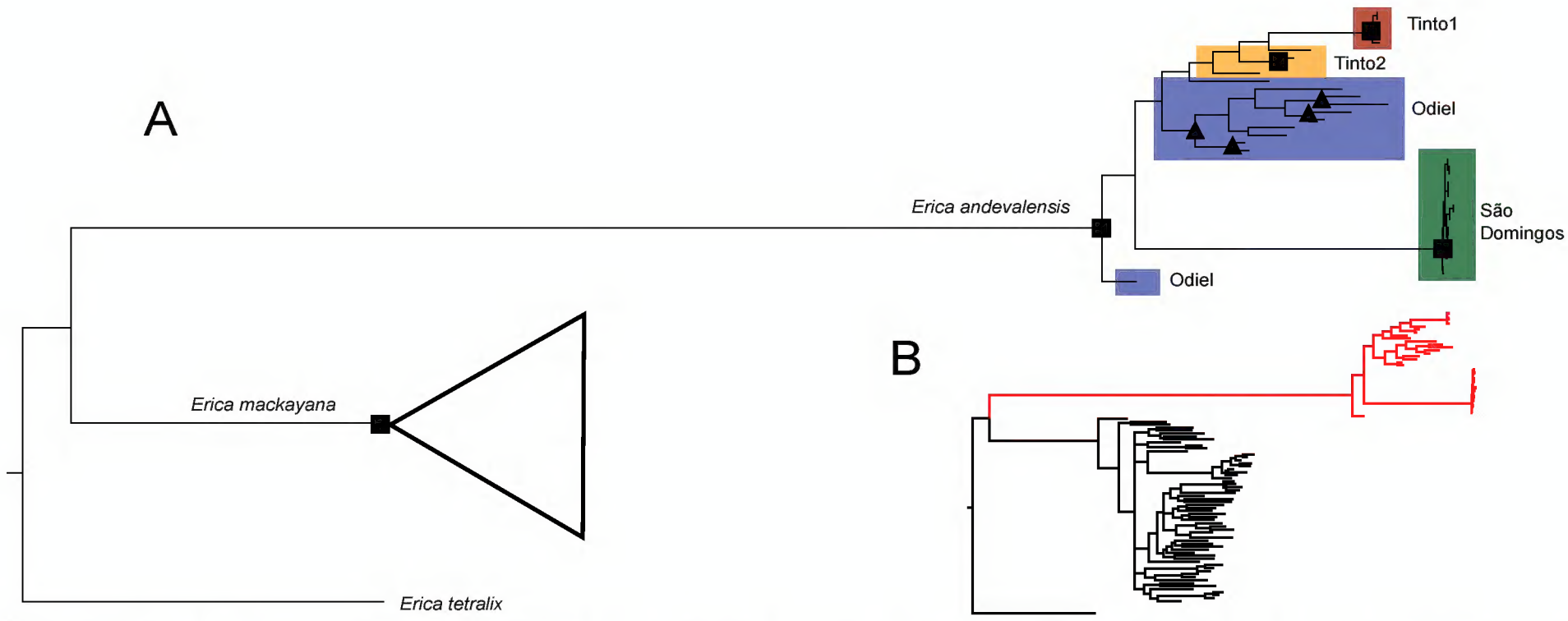

Figure 6. Phylogenetic tree of *E. andevalensis* and *E. mackayana* **A** tree showing the *E. mackayana* clade collapsed. Colors represent sites as in Fig. 4. Bootstrap support is indicated on nodes when >80: squares represent full support and triangles 80–99% **B** expanded tree.

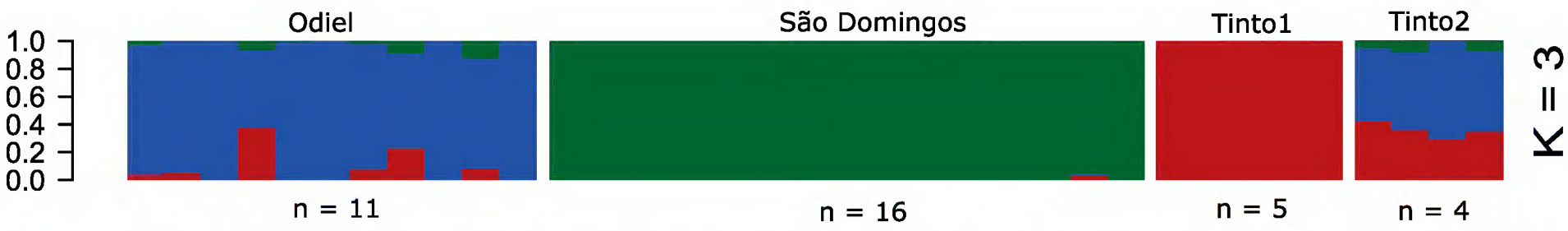

Figure 7. Bayesian individual assignment probabilities of all populations belonging to *E. andevalensis*. K = 3, was calculated by cross-entropy scores.

Populations structure of *E. andevalensis*

Bayesian population assignment probabilities revealed a highly structured assemblage into three clusters in *E. andevalensis*. This number of clusters was the optimal aggrupation within the dataset following cross-entropy scores. The assemblage is consistent with geographical distribution of populations from São Domingos and Tinto1, which were the most homogeneous, and admixture with other clusters is nearly absent. These two populations corresponded with two of the clusters (green and red, respectively in Fig. 7). The third cluster (blue) was dominant in the population from Odiel, in which the other two clusters were also represented. Finally, all plants from Tinto2 population showed high levels of admixture mainly from red and blue clusters.

We also assessed population structure through pairwise F_{ST} comparisons. The results align with clustering analyses, providing additional insights. Notably, São Domingos stood out as a genetically distinct population (F_{ST} = 0.906; 0.8; 0.652, from Tinto1, Tinto2 and Odiel populations respectively). Spanish populations (Odiel, Tinto1, and Tinto2) showed a lower genetic differentiation, although F_{ST} values among populations were relatively high (Odiel-Tinto1, F_{ST} = 0.392; Tinto2-Tinto1: 0.466). Only the genetic differentiation between Odiel and Tinto2 was moderate low (Fst = 0.122).

The nucleotide diversity (π_i) and heterozygosity (H_o and H_e) in *E. andevalensis* exhibited significant variation among populations, with Odiel and Tinto2 ranking as the most diverse, in that order. Additionally, these two populations showed a lower F_{IS} index (Table 2). In contrast, São Domingos displayed an

Table 2. Summary diversity statistics of *E. andevalensis* populations calculated across sampling locations in which more than two samples were collected.

Population	N	H_o	H_e	F_{IS}	π_i
Odiel	11	0.155	0.289	0.464	0.219
São Domingos	16	0.028	0.087	0.674	0.025
Tinto 1	5	0.049	0.106	0.537	0.044
Tinto 2	4	0.148	0.247	0.401	0.163
Total	**36**	**0.083**	**0.169**	**0.508**	**0.10**

exceptionally low level of diversity, being more than eight times less nucleotide diverse than the Odiel population. Furthermore, São Domingos emerged as the population with a higher excess of homozygotes than expected (F_{IS} = 0.674).

Nucleotide diversity (π_i) and heterozygosity (H_o and H_e) in *E. andevalensis* was in general similar to *E. mackayana* (Suppl. material 1: table S6). However, Odiel and Tinto2 populations were more diverse than those of *E. mackayana*, while Tinto1 and São Domingos are less. F_{IS} index was high in both species for the two species in all sampling areas, but higher in *E. andevalensis*.

Discussion

This study provides evidence supporting the status of *Erica andevalensis* as an accepted species, clearly distinct from its sister-group *E. mackayana*. Both morphological and genetic variability between species is much higher than among populations, and populations are highly structured in *E. andevalensis*, as previously found for *E. mackayana* (Fagúndez and Díaz-Tapia 2023). *Erica andevalensis* and *E. mackayana* share a number of traits that are rare or unique in the northern heathers such as broad leaves, umbel-like terminal inflorescences and the presence of pluricellular glandular hairs in leaves and stems. Together with *E. tetralix* and *E. ciliaris*, they belong to a robust clade that exhibits further unique features such as pluricellular glandular indumentum or unrolled leaf margins, compared to other paleartic species (Bayer 1993; Nelson 2012; Mugrabi de Kuppler et al. 2015).

Morphology and phenotypic analysis

The analysis of morphological traits provided reliable diagnostic characters for *E. mackayana* and *E. andevalensis* (Table 1, Figs 4, 5). These included the presence of some short hairs in the apex of the ovary near the insertion point of the style in *E. andevalensis*, which are absent in *E. mackayana*. This was previously stated by Bayer (1993), but not in the original description of *E. andevalensis*, in which it is described as having a glabrous ovary (Cabezudo and Rivera 1980). Variation in ovary hairiness is one of the most informative traits among species of the northern heathers, varying consistently between species even within closely related groups such as the *E. ciliaris-tetralix* clade (Nelson 2012). Hybridization among species of this group is recognized by the presence of hairs in the ovary, generally fewer and localized towards the apex as in *E.* × *stuartii* (*E. tetralix* × *E. mackayana*, Fagúndez 2006).

Other diagnostic characters related to leaf indumentum included the presence of short unicellular hairs in *E. andevalensis* which are nearly absent in *E. mackayana*, and long, pluricellular glandular trichomes in the two species but

much longer in *E. mackayana*. Diagnostic differences in leaf hairiness are also reflected in the hairiness of other vegetative and non-vegetative organs, such as stems and flower pedicels (Cabezudo and Rivera 1980).

Leaf arrangement also contributed to species delimitation, as *E andevalensis* consistently shows 4-nate whorls of erect leaves. In *E. mackayana*, leaves are 4-6-nate and patent. Leaves of *E. andevalensis* are similar in size to those of *E. mackayana*, but narrower with more rolled-in margins. Remarkably, leaf length shows minimum contribution to PC1, but the highest weight in PC2, reflecting strong variation at the population level (Suppl. material 1: table S4). Leaf size is commonly described as a plant trait with high variation due to phenotypic plasticity (Stotz et al. 2022). This character may be informative to understand adaptation to stress factors or environmental constraints such as climate and soil condition in this group of heathers.

Other traits showed limited potential as diagnostic characters, particularly those related to reproductive biology. In *E. andevalensis*, flowers can be shorter and narrower, with a more pronounced style exertion. However, flower size and proportions can vary at different development stages. This variability is even more pronounced in ovary morphology and size (Suppl. material 1: table S3). *E. andevalensis* tends to have larger but flatter ovaries, but ovary size and shape depends on its development stage in transition towards fruit formation, which is difficult to assess. With regards to anther traits, *E. andevalensis* has darker and longer anthers, with a wider aperture, and a larger knob (a small protuberance in the anther) than *E. mackayana*, but these traits showed high overlapping values (Suppl. material 1: table S3).

Genetic identity of *Erica andevalensis*

Phylogenetic analyses show that both *Erica mackayana* and *E. andevalensis* are well supported monophyletic groups. Reciprocal monophyly is accepted as one of the most important lines of evidence on species delimitation (Moritz 1994; Mehta et al. 2019). Genetic distance has also been used extensively in species delimitation. We found that genetic distance was >4-fold higher between the species *E. andevalensis* and *E. mackayana* than among their populations. A much greater distance between species compared to populations supports their consideration as different species, in line with Birky et al. (2010) and Dellicour and Flot (2015). A combined framework for species delimitation, incorporating both phylogenetic tree-based approaches and genetic distance analysis, has been applied across various taxonomic groups (e. g. Bradley and Baker 2001; Meudt et al. 2009; Del Prado et al. 2011; Goicoechea et al. 2012).

The much greater branch lengths in *E. andevalensis* than in *E. mackayana* (Fig. 6), are consistent with results from phylogenetic analysis of plastid and nuclear ribosomal DNA sequences in Mugrabi de Kuppler et al. (2015) and could reflect a consistently faster rate of sequence divergence in *E. andevalensis*. Smaller population sizes resulting from edaphic specialization over time could contribute to this phenomenon. Small population size has been correlated with fast evolution empirically in many groups (Lanfear et al. 2014). There is potential for further exploration of the relation between genomic processes and enhanced ability of *E. andevalensis* to adapt to extreme environments. *Erica andevalensis*, has been extensively studied for its metal-tolerant characteristics, thriving in heavily polluted soils

with high concentrations of Cu, Ni, or F (e.g. Márquez et al. 2005; Rossini-Oliva et al. 2018). Its close relative, *E. mackayana*, can grow in a wider range of soils, even in serpentine soils with high levels of heavy metals, (e.g. A Capelada mountain range in NW Spain) and can inhabit areas with pH as low as 3 (Webb 1955; Fagúndez and Pontevedra-Pombal 2022), so may also exhibit pyritic-soil tolerance.

Intraspecific variation in *Erica andevalensis*

Two of the studied populations of *E. andevalensis* showed low support for internal nodes in the phylogeny, but the population structure analysis showed low levels of admixture, meaning that *E. andevalensis* populations are highly structured. The genetic differentiation among populations separated by less than 85 km is higher than in the entire distribution area analyzed for the related species *E. mackayana* (Fagúndez and Díaz-Tapia 2023). A high differentiation among populations is typically found in taxa with a highly fragmented distribution, as in soil endemic plants (Leimu and Mutikainen 2005; Nistelberger et al. 2015). The large genetic divergence between Tinto1 and Tinto2, separated by only 7.5 km and initially considered as a single population, was a surprising finding. We found a lower proximity due to river section (medium *vs* low river course) than that of river identity (Odiel *vs* Tinto).

Soil endemic plants usually have low genetic diversity, but their F_{IS} index clearly diverge between different plants groups (e. g. Barbará et al. 2007; Wang et al. 2017; Nagasawa et al. 2019). It has been suggested that species traits such as self-compatibility or clonality might impact the F_{IS} index (Lavor et al. 2014), thus comparing soil endemic species F_{IS} values with related taxa with broader ranges can be interesting. Compared to its close relative *E. mackayana*, *E. andevalensis* had higher F_{IS} index values, and genetic diversity was highly variable among *E. andevalensis* populations. São Domingos population results show the highest F_{IS} value and lowest genetic diversity of all populations of both species, in line with results from the *E. andevalensis* microsatellites' analyses (Bandeira de Albuquerque et al. 2008), and probably reflecting a founder effect. This population is at the edge of the species distribution range, entirely located in an abandoned mining area (Figs 1C, 2).

The origin and migration history of different *Erica* species are a subject of debate, especially when there is a potential transportation linked to human activities (Fagúndez and Díaz-Tapia 2023; Skeffington and Scott 2023). Nelson et al. (1985) suggested that mining populations of *E. andevalensis* might not be native, originating with the commencement of mining activity in the area, estimated by Tornos et al. (2000) to be more than 5000 years ago. This hypothesis aligns with the low genetic diversity of São Domingos, but Bandeira de Albuquerque et al. (2008) found high genetic diversity in other populations in Huelva linked to mining activities. Nevertheless, the creation of new niches by humans (highly toxic damp mines), coupled with an increased concentration of heavy metals in rivers, could have facilitated the establishment of larger populations beyond their natural limits and possibly a recent migration between previously unconnected populations. This is consistent with the moderate connection between populations in the lower Odiel and Tinto basins and could explain higher genetic diversity in these populations and slightly lower F_{IS} index. The small size of the seeds which are produced in large numbers makes them easily dispersed by animals, wind and other vectors including humans (Aparicio and García-Martín 1996; Fagúndez and Izco 2004; Fagúndez and Izco 2010).

Implications for conservation

Clarifying cryptic or poorly understood taxa and species delimitation is needed for the design of conservation policies and actions, especially in large genera such as *Erica* (Pirie et al. 2022). Verification of specific status should give stronger support to the natural value of the Iberian Pyrite Belt, which is partly a protected area in Spain (BOJA 2005). *Erica andevalensis* is legally considered an endangered species, classified as vulnerable (VU), but only in Spain. The low genetic diversity and singularity of the Portuguese population, both phenotypical and genotypical, coupled with its presence in a very restricted area, should justify the consideration of *E. andevalensis* as an endangered species in Portugal as well, as proposed in Carapeto et al. (2020).

Conservation strategies for *E. andevalensis* should protect a variety of populations in different river basins and sections, and research evaluating its use for restoration should consider infraspecific variability and special issues of working with metallophyte species. For instance, habitat decontamination of heavy metal polluted river shores needs to be carefully considered, as it may have a negative effect in local populations of *E. andevalensis* (Márquez et al. 2005). Recommended conservation actions of metallophytes endemic species usually include prohibiting new mining activities in areas (Whiting et al. 2004). This recommendation was also made to conserve *Erica andevalensis* populations (Márquez et al. 2005). However, certain metallophytes may benefit from human perturbations in the long-term, potentially leading to an increase in their populations (Faucon et al. 2011), and this may be the case for *E. andevalensis*. Many conservation studies on metallophytes focus on areas where mining has been intensive but only during the last century. Our study, by contrast, shows the genetic and dispersal effects of long-term mining (>5000 years BP) in an endemic metallophyte species.

There is potential for use of *E. andevalensis* in bioremediation of polluted soils (Abreu et al. 2008; Monaci et al. 2010; Pérez-López et al. 2014). Differences with regards to soil preference between Spanish and Portuguese populations (Pérez-López et al. 2014), suggest a potential variability in bioremediation performance at the population level. These ecological differences, also reflected in genetics, should be considered in planning habitat restoration.

Acknowledgements

We would like to thank M. Sheehy Skeffington, R. Sheppard and A. Montañés for field support. We also acknowledge the computational facilities of Centro de Supercomputación de Galicia (CESGA). APCs were covered by CICA (Centro Interdisciplinar de Química e Bioloxía), UDC.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

No funding was reported.

Author contributions

All authors have contributed equally.

Author ORCIDs

Iván Rodríguez-Buján https://orcid.org/0000-0002-9869-4033
Pilar Díaz-Tapia https://orcid.org/0000-0003-4680-4867
Jaime Fagúndez https://orcid.org/0000-0001-6605-7278

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Abreu MM, Tavares MT, Batista MJ (2008) Potential use of *Erica andevalensis* and *Erica australis* in phytoremediation of sulphide mine environments: São Domingos, Portugal. Journal of Geochemical Exploration 96(2–3): 210–222. https://doi.org/10.1016/j.gexplo.2007.04.007

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews. Genetics 17(2): 81–92. https://doi.org/10.1038/nrg.2015.28

Aparicio A (1999) *Erica andevalensis*. In: Valdés Castrillón C, Rodríguez Hiraldo A, López Ontiveros O, Merino O (Eds) Libro Rojo de la Flora Silvestre Amenazada de Andalucía. Tomo I: Especies en peligro de extinción. Consejería de Medio Ambiente. Junta de Andalucía, 119–122. https://www.biolveg.uma.es/links/libro_rojo_andalucia_tomo_I.pdf

Aparicio A, García-Martín F (1996) The reproductive biology and breeding system of *Erica andevalensis* Cabezudo & Rivera (Ericaceae), an endangered edaphic endemic of southwestern Spain. Implications for its conservation. Flora (Jena) 191(4): 345–351. https://doi.org/10.1016/S0367-2530(17)30739-9

Bandeira de Albuquerque M, Rodríguez-Echeverría S, Freitas H (2008) Genetic diversity in populations of *Erica andevalensis*, a vulnerable metallophyte species from the Iberian Peninsula. Web Ecology 8(1): 135–141. https://doi.org/10.5194/we-8-135-2008

Barbará T, Martinelli G, Fay MF, Mayo SJ, Lexer C (2007) Population differentiation and species cohesion in two closely related plants adapted to neotropical high-altitude 'inselbergs', *Alcantarea imperialis* and *Alcantarea geniculata* (Bromeliaceae). Molecular Ecology 16(10): 1981–1992. https://doi.org/10.1111/j.1365-294X.2007.03272.x

Bayer E (1993) *Erica*. In: Castroviejo S, Aedo C, Gómez Campo C, Laínz M, Montserrat P, Morales R, Muñoz Garmendia F, Nieto Feliner G, Rico E, Talavera S, Villar L (Eds) Flora iberica. Plantas vasculares de la Península Ibérica e Islas Baleares. Vol IV. Cruciferae-Monotropaceae. Real Jardín Botánico, Madrid, 484–505.

Birky Jr CW, Adams J, Gemmel M, Perry J (2010) Using population genetic theory and DNA sequences for species detection and identification in asexual organisms. PLOS ONE 5(5): e10609. https://doi.org/10.1371/journal.pone.0010609

BOJA (2005) Decreto 558/2004, de 14 de diciembre, por el que se declara el Paisaje Protegido de Río Tinto. Boletín de la Junta de Andalucía.

BOJA (2012) Decreto 23/2012, de 14 de febrero, por el que se regula la conservación y el uso sostenible de la flora y la fauna silvestres y sus hábitats. Boletín de la Junta de Andalucía.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics (Oxford, England) 23(19): 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bradley RD, Baker RJ (2001) Test of the genetic species concept: Cytochrome-b sequences and mammals. Journal of Mammalogy 82(4): 960–973. https://doi.org/10.1644/1545-1542(2001)082%3C0960:ATOTGS%3E2.0.CO;2

Buira A, Carapeto A, García-Murillo PG, Monteiro-Henriques T (2017) *Erica andevalensis*. The IUCN Red List of Threatened Species 2017: e.T102819769A102819777. https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T102819769A102819777.en

Cabezudo B, Rivera J (1980) Notas taxonómicas y corológicas sobre la Flora de Andalucía Occidental 2: *Erica andevalensis* Cabezudo & Rivera sp. nov. Lagascalia 9(2): 223–226.

Carapeto A, Francisco A, Pereira P, Porto M (2020) Sociedade Portuguesa de Botânica e Associação Portuguesa de Ciência da Vegetação. Phytos, 372 pp.

Coates DJ, Byrne M, Moritz C (2018) Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. Frontiers in Ecology and Evolution 6: 165. https://doi.org/10.3389/fevo.2018.00165

De Jong MJ, de Jong JF, Hoelzel AR, Janke A (2021) SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. Molecular Ecology Resources 21(4): 1369–1379. https://doi.org/10.1111/1755-0998.13339

De Queiroz K (2005) Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America 102(suppl_1): 6600–6607. https://doi.org/10.1073/pnas.0502030102

De Queiroz K (2007) Species concepts and species delimitation. Systematic Biology 56(6): 876–886. https://doi.org/10.1080/10635150701701083

Del Prado R, Divakar PK, Crespo A (2011) Using genetic distances in addition to ITS molecular phylogeny to identify potential species in the *Parmotrema reticulatum* complex: A case study. Lichenologist (London, England) 43(6): 569–583. https://doi.org/10.1017/S0024282911000582

Dellicour S, Flot JF (2015) Delimiting species-poor data sets using single molecular markers: A study of barcode gaps, haplowebs and GMYC. Systematic Biology 64(6): 900–908. https://doi.org/10.1093/sysbio/syu130

Elliott A, Bester SP, Klopper RR, Nelson EC, Pirie M (2024) Curating an online checklist for *Erica* L. (Ericaceae): contributing to and supporting global conservation through the World Flora Online. ARPHA Preprints 5: e121772. https://doi.org/10.3897/arpha-preprints.e121772

Fagúndez J (2006) Two wild hybrids of *Erica* L. (Ericaceae) from northwest Spain. Botanica Complutensis 30: 131.

Fagúndez J (2013) Heathlands confronting global change: Drivers of biodiversity loss from past to future scenarios. Annals of Botany 111(2): 151–172. https://doi.org/10.1093/aob/mcs257

Fagúndez J, Díaz-Tapia P (2023) Comparative phylogeography of a restricted and a widespread heather: Genetic evidence of multiple independent introductions of *Erica mackayana* into Ireland from northern Spain. Botanical Journal of the Linnean Society 201(3): 329–340. https://doi.org/10.1093/botlinnean/boac071

Fagúndez J, Izco J (2004) Taxonomic value of seed characters in the *Erica tetralix* L. group (Ericaceae). Plant Biosystems 138(3): 207–213. https://doi.org/10.1080/11263500400021059

Fagúndez J, Izco J (2010) Seed morphology of the European species of *Erica* L. sect. *Arsace* Salisb. ex Benth. (Ericaceae). Acta Botanica Gallica 157(1): 45–54. https://doi.org/10.1080/12538078.2010.10516188

Fagúndez J, Pontevedra-Pombal X (2022) Soil properties of North Iberian wet heathlands in relation to climate, management and plant community. Plant and Soil 475(1–2): 565–580. https://doi.org/10.1007/s11104-022-05393-6

Faucon MP, Parmentier I, Colinet G, Mahy G, Ngongo Luhembwe M, Meerts P (2011) May rare metallophytes benefit from disturbed soils following mining activity? The case of the *Crepidorhopalon tenuis* in Katanga (DR Congo). Restoration Ecology 19(3): 333–343. https://doi.org/10.1111/j.1526-100X.2009.00585.x

Feng JY, Li M, Zhao S, Zhang C, Yang ST, Qiao S, Tan WF, Qu HJ, Wang DY, Pu ZG (2018) Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species *I. trifida* using RAD-seq. BMC Plant Biology 18(1): 1–12. https://doi.org/10.1186/s12870-018-1399-x

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143(9): 1919–1927. https://doi.org/10.1016/j.biocon.2010.05.011

Galtier N (2019) Delineating species in the speciation continuum: A proposal. Evolutionary Applications 12(4): 657–663. https://doi.org/10.1111/eva.12748

GBIF (2023) Global Biodiversity Information Facility. [Published on the internet:] https://www.gbif.org [Accessed 30.10.2023]

Goicoechea N, Padial JM, Chaparro JC, Castroviejo-Fisher S, De la Riva I (2012) Molecular phylogenetics, species diversity, and biogeography of the Andean lizards of the genus *Proctoporus* (Squamata: Gymnophthalmidae). Molecular Phylogenetics and Evolution 65(3): 953–964. https://doi.org/10.1016/j.ympev.2012.08.017

Kalyaanamoorthy S, Minh B, Wong T, von Haeseler A, Jeermin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods 14(6): 587–589. https://doi.org/10.1038/nmeth.4285

Kassambara A, Mundt F (2020) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=-factoextra

Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. Trends in Ecology & Evolution 29(1): 33–41. https://doi.org/10.1016/j.tree.2013.09.009

Lavor P, van den Berg C, Jacobi CM, Carmo FF, Versieux LM (2014) Population genetics of the endemic and endangered *Vriesea minarum* (Bromeliaceae) in the Iron Quadrangle, Espinhaço Range, Brazil. American Journal of Botany 101(7): 1167–1175. https://doi.org/10.3732/ajb.1300388

Leimu R, Mutikainen P (2005) Population history, mating system, and fitness variation in a perennial herb with a fragmented distribution. Conservation Biology 19(2): 349–356. https://doi.org/10.1111/j.1523-1739.2005.00480.x

Manning J, Goldblatt P (2012) Plants of the Greater Cape Floristic Region 1: the Core Cape flora, Strelitzia 29. South African National Biodiversity Institute, Pretoria.

Márquez B, Hidalgo PJ, Heras MA, Velasco R, Córdoba F (2005) *Erica andevalensis*: Un brezo endémico y en peligro de extinción de la zona minera de huelva. Jornadas Técnicas de Ciencias Ambientales, 1–19.

McClintock DC, Nelson EC (1989) The heathers of Europe and adjacent areas. Botanical Journal of the Linnean Society 101(3): 279–289. https://doi.org/10.1111/j.1095-8339.1989.tb00162.x

Mehta RS, Rosenberg NA (2019) The probability of reciprocal monophyly of gene lineages in three and four species. Theoretical Population Biology 129: 133–147. https://doi.org/10.1016/j.tpb.2018.04.004

Meudt HM, Lockhart PJ, Bryant D (2009) Species delimitation and phylogeny of a New Zealand plant species radiation. BMC Evolutionary Biology 9(1): 111. https://doi.org/10.1186/1471-2148-9-111

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37(5): 1530–1534. https://doi.org/10.1093/molbev/msaa015

Monaci F, Leidi EO, Mingorance MD, Valdes B, Rossini Oliva S, Bargagli R (2010) Phytostabilization and re-vegetation of mine tailings at Riotinto (SW Spain): the potential of two spontaneous shrubs (*Erica andevalensis* and *E. australis*). In International Conference on Environmental Pollution and Clean Bio/Phytoremediation.

Moritz C (1994) Defining 'evolutionarily significant units' for conservation. Trends in Ecology & Evolution 9(10): 373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Mugrabi de Kuppler AL, Fagúndez J, Bellstedt DU, Oliver EG, Léon J, Pirie MD (2015) Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). Molecular Phylogenetics and Evolution 88: 121–131. https://doi.org/10.1016/j.ympev.2015.04.005

Nagasawa K, Setoguchi H, Maki M, Goto H, Fukushima K, Isagi Y, Suyama Y, Matsuo A, Tsunamoto Y, Sawa K, Sakaguchi S (2019) Genetic consequences of plant edaphic specialisation to solfatara fields; phylogenetic and population genetic analysis of *Carex angustisquama* (Cyperaceae). Molecular Ecology 29(17): 3234–3247. https://doi.org/10.1111/mec.15324

Nelson EC (2012) Hardy Heathers from the Northern Hemisphere: *Calluna*, *Daboecia*, *Erica*. Kew Publishing, Royal Botanic Gardens, Kew.

Nelson EC, Small DJ (2000) International register of heather names. Volume 1: hardy cultivars and European species, (A-Z). The Heather Society.

Nelson EC, McClintock DC, Small D (1985) The natural habitat of *Erica andevalensis* in south-western Spain. Curtis's Botanical Magazine 2(3): 324–330. https://doi.org/10.1111/j.1467-8748.1985.tb00264.x

Nistelberger HM, Byrne M, Coates D, Roberts JD (2015) Phylogeography and population differentiation in terrestrial island populations of *Banksia arborea* (Proteaceae). Biological Journal of the Linnean Society, Linnean Society of London 114(4): 860–872. https://doi.org/10.1111/bij.12464

Oliver EGH, Forshaw N, Oliver IM, Volk F, Schumann AWS, Dorr LJ, Hoekstra RD, Musker SD, Nürk NM, Pirie MD, Rebelo AG (2024) Genus *Erica*: An identification aid version 4.00. PhytoKeys 241: 143–154. https://doi.org/10.3897/phytokeys.241.117604

Pante E, Schoelinck C, Puillandre N (2015) From integrative taxonomy to species description: One step beyond. Systematic Biology 64(1): 152–160. https://doi.org/10.1093/sysbio/syu083

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: A road map for stacks. Methods in Ecology and Evolution 8(10): 1360–1373. https://doi.org/10.1111/2041-210X.12775

Pérez-López R, Márquez-García B, Abreu MM, Nieto JM, Córdoba F (2014) *Erica andevalensis* and *Erica australis* growing in the same extreme environments: Phytostabilization potential of mining areas. Geoderma 230: 194–203. https://doi.org/10.1016/j.geoderma.2014.04.004

Pirie MD, Blackhall-Miles R, Bourke G, Crowley D, Ebrahim I, Forest F, Knaack M, Koopman R, Lansdowne A, Nürk NM, Osborne J, Pearce TR, Rohrauer D, Smit M, Wilman V (2022) Preventing species extinctions: A global conservation consortium for *Erica*. Plants, People, Planet 4(4): 335–344. https://doi.org/10.1002/ppp3.10266

Pirie M, Bellstedt D, Bouman R, Fagúndez J, Gehrke B, Kandziora M, Maitre NCL, Musker S, Newman E, Nürk N, Oliver EGH, Pipins S, van der Niet T, Forest F (2024) Spatial decoupling of taxon richness, phylogenetic diversity, and threat status in the megagenus *Erica*. ARPHA Preprints 5: e124629. https://doi.org/10.3897/arphapreprints.e124629

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rossini-Oliva S, Abreu M, Leidi EO (2018) A review of hazardous elements tolerance in a metallophyte model species: *Erica andevalensis*. Geoderma 319: 43–51. https://doi.org/10.1016/j.geoderma.2017.12.035

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9(7): 671–675. https://doi.org/10.1038/nmeth.2089

Skeffington MS, Scott N (2023) Were the five rare heathers of the west of Ireland introduced through human activity? An ecological, genetic, biogeographical and historical assessment. British & Irish Botany 5(2). https://doi.org/10.33928/bib.2023.05.221

Šlenker M, Koutecký P, Marhold K (2022) MorphoTools2: An R package for multivariate morphometric analysis. Bioinformatics (Oxford, England) 38(10): 2954–2955. https://doi.org/10.1093/bioinformatics/btac173

Soltis PS, Soltis DE (2012) Polyploidy and genome evolution (Vol. 665). Springer, Berlin. https://doi.org/10.1007/978-3-642-31442-1

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods a bioconductor package providing PCA methods for incomplete data. Bioinformatics (Oxford, England) 23(9): 1164–1167. https://doi.org/10.1093/bioinformatics/btm069

Stotz GC, Salgado-Luarte C, Escobedo VM, Valladares F, Gianoli E (2022) Phenotypic plasticity and the leaf economics spectrum: Plasticity is positively associated with specific leaf area. Oikos 2022(11): e09342. https://doi.org/10.1111/oik.09342

Szkudlarz P (2009) Variation in seed morphology in the genus *Erica* L. (Ericaceae). Biodiversity Research and Conservation 16(2009): 1–106. https://doi.org/10.2478/v10119-009-0022-2

Tamura K, Nei N (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10(3): 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biology and Evolution 38(7): 3022–3027. https://doi.org/10.1093/molbev/msab120

Torgo L (2016) Data Mining with R, learning with case studies, 2nd edn. Chapman and Hall/CRC. https://doi.org/10.1201/9781315399102

Tornos F, Barriga F, Marcoux E, Pascual E, Pons JM, Relvas J, Velasco F (2000) The Iberian pyrite belt. Database on Global VMS Districts, Codes-Geode, 19–52.

Wang J, Feng C, Jiao T, Von Wettberg EB, Kang M (2017) Genomic signature of adaptive divergence despite strong nonadaptive forces on edaphic islands: a case study

of *Primulina juliae*. Genome Biology and Evolution 9(12): 3495–3508. https://doi.org/10.1093/gbe/evx263

Webb DA (1955) Biological flora of the British Isles. *Erica mackayana*. Journal of Ecology 43: 319–330. https://doi.org/10.2307/2257145

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. Evolution; International Journal of Organic Evolution 38(6): 1358–1370. https://doi.org/10.2307/2408641

WFO (2024) World Flora Online. [Published on the Internet:] http://www.worldfloraonline.org [Accessed 04.04.2024]

Whiting SN, Reeves RD, Richards D, Johnson MS, Cooke JA, Malaisse F, Paton A, Smith JAC, Angle JS, Chaney RL, Ginocchio R, Jaffré T, Johns R, McIntyre T, Purvis OW, Salt DE, Schat H, Zhao FJ, Baker AJM (2004) Research priorities for conservation of metallophyte biodiversity and their potential for restoration and site remediation. Restoration Ecology 12(1): 106–116. https://doi.org/10.1111/j.1061-2971.2004.00367.x

Whittemore AT (1993) Species concepts: A reply to Ernst Mayr. Taxon 42(3): 573–583. https://doi.org/10.2307/1222535

Supplementary material 1

Supplementary tables

Authors: Iván Rodríguez-Buján, Pilar Díaz-Tapia, Jaime Fagúndez

Data type: docx

Explanation note: **table S1:** Information on the populations of the studied species *Erica mackayana* and *E. andevalensis*, and one population of *E. tetralix* used as the outgroup. N_1: number of individuals used in the genetic analyses, based on having <50% of missing data in population structure analysis. N_2: number of individuals used in the morphological analyses. **table S2:** Morphological traits measured in plants of *Erica mackayana* and *E. andevalensis*, acronym names and methodology. QN = Quantitative trait, O = Ordinal trait, B = Binary trait. N = Number of measures per specimen. **table S3:** Results of univariant tests between *Erica mackayana* and *E. andevalensis*, including the arithmetic mean and range for each of the morphological traits named as in Fig. 1 and Suppl. material 1: table S2. *E. mack* = *E. mackayana*. *E. andev*. = *E. andevalensis*. * = p-value < 0.05, **= p-value < 0.01, *** = p-value < 0.001. **table S4:** Percentage of contribution of each morphological trait included in the PCA to the first two components (PC1 and PC2), named as in Fig. 1 and Suppl. material 1: table S2, ordered by their contribution to PC1. **table S5:** Distance matrix for the studied populations of *Erica mackayana* and *E. andevalensis*. Pairwise Tamura and Nei distance (1993) calculated upon MEGA 11. **table S6:** Summary of the genetic diversity statistics of *E. andevalensis* and *E. mackayana* populations. Statistics for populations with less than three individuals were not included.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/phytokeys.244.120914.suppl1